Spring 2015

SHARP RNA Recognition Motif Optimizations, Extensions, and Mutations for Use in 2D and 3D NMR Experiments

Shaun M. Christie
University of Akron Main Campus, smc148@zips.uakron.edu

Title: SHARP RNA Recognition Motif Optimizations, Extensions, and Mutations for Use in 2D and 3D NMR Experiments

Name: Shaun Christie

Course: 3150:497

Date: 4/27/2015

SHARP RNA Recognition Motif Optimizations, Extensions, and Mutations for Use in 2D and 3D NMR Experiments

Shaun Christie

University of Akron

Abstract: SMRT/HDAC Associated Repressor Protein interacts with the long noncoding RNA, produced by SRA, by binding at the RRMs. Three projects were done to prepare the truncated proteins for use in 2D and 3D NMR experiments. The first focuses on RRM 3 and its optimization during the purification process. The second focuses on RRM 2-4, which was found to be missing two alpha helices that may be important for protein stability. These helices can also interact with RRM 3 as well due to the tight association of RRMs 3 and 4. The two step PCR extension of RRM 2-4 was assumed to work based on sequencing and agarose gels. The final project focused on serine to cysteine mutations of RRM 2 to allow site directed spin labeling for use in NMR PRE experiments. Both mutants developed in this project do not affect the protein structure or the residues surrounding the mutation based on wild type and mutant HSQC spectra. The spectra of the spin labeled samples does show broadening some resonances, but may need reductant removed from the buffer after the labeling reaction.

Abbreviations: PDB (protein data bank, URL: rcsb.org). PRE (paramagnetic relaxation enhancement). NOE (nuclear Overhauser effect). RRM (RNA recognition motif). SHARP (SMRT/HDAC Associated Repressor Protein). SRA (steroid receptor activator gene). IMAC (immobilized metal affinity chromatography).

Introduction

The SMRT/HDAC Associated Repressor Protein (SHARP) contains four RNA binding domains that are believed to interact with the long non-coding RNA produced by the steroid receptor activator

gene (SRA). These RNA recognition motifs (RRMs) are specific and necessary for protein-RNA binding since deletions of the RRMs have shown a loss of interaction. SHARP and SRA RNA are involved in the epigenetic regulation of transcription, Notch signaling, and various cancers[1]. Previous experiments in Dr. Leeper's lab had determined that the isolation of the protein containing all four RRMs was difficult under normal purification conditions. In order to study the motifs individually, site directed mutagenesis (SDM) was used to insert a stop codon at residues between each motif in the parent plasmid resulting in plasmid production for the RRMs 1, 2, 3, 4, 2-3, and 2-4. RRMs 1 and 2 have been successfully purified and concentrated for use in nuclear magnetic resonance spectroscopy (NMR). The other plasmids produce proteins at very low yields and only if the protein did not precipitate during the purification process. 2D and 3D NMR experiments are used to study biomacromolecular structures and complexes, however, these experiments require protein samples with concentrations of at least 0.2 mM in a volume of 0.5 mL or more. NMR is used to determine angles, chemical shifts, and atomic distances of a macromolecule in solution. When the protein is in solution it allows observation of the complex behavior and kinetics in real time rather than the static conformation observed by using X-ray crystallography. Multiple experiments are run to determine backbone and side chain resonances allowing a structure to be modeled from this data. The previous attempts to concentrate RRM 3 and RRM 2-4 (containing three of the four RRMs) resulted in aggregation of final samples or concentrations of no greater than 0.1 mM.

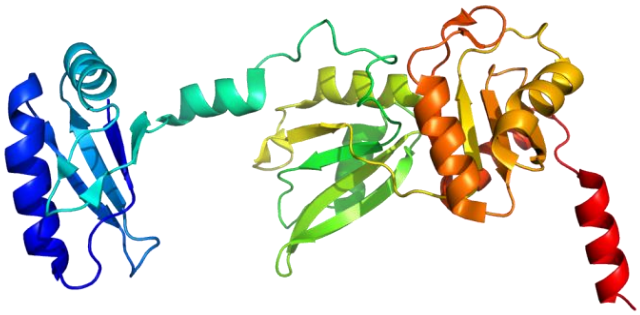


Figure 1- SHARP RRM 2-4 x-ray crystallography structure (PDB 4P6Q). RRM 2 (blue), RRM 3 (green), and RRM 4 (orange).

For the purification of RRM 3, multiple pH buffer conditions had been tested by students in Dr. Leeper's group. Aggregation occurs due to the shift in the isoelectric point (pI) after cleavage of the 6-Histidine tag used during immobilized metal affinity chromatography (IMAC) purification. These tests determined that a pH 6 buffer caused the least aggregation of protein. However, the first project described here attempted to increase the concentration of the final RRM 3 product. Multiple buffer conditions were tested based on the review *Protein Production and Purification*[2].Conditions included reductant concentration, salt type, salt concentration, and addition of amino acids. The primary reductant used in the buffers is dithiothreitol (DTT), which prevents formation of disulfide bridges between cysteine residues of the protein. RRM 3 contains three cysteine residues, two on the surface facing solution and one in the unstructured C-terminal region. Changes in salt type and concentration were used to determine whether ionic interactions cause aggregation. The addition of arginine and glutamate to the purification buffers deters the protein-protein aggregation by interactions with hydrophobic patches on the protein surface, but the exact mechanism is currently unknown. The addition of arginine and glutamate does not adversely affect solution NMR or protein-RNA complexes[3].

Arieti et al. (2014)[4] were able to purify the SHARP 2-4 protein for use in X-ray crystallography experiments. Observation of the group's structure (Figure 1) determined that the SHARP 2-4 plasmid used by Dr. Leeper's lab was missing a portion of the protein's C-terminal end. These 27 residues comprise two alpha helices that may be involved in stabilizing the protein (Appendix 2: Figure 1). The first helix (residues 588-602) covers the beta sheet of RRM 4, which contains the proposed RNA binding site. However, other proteins exhibiting this feature undergo a conformational change that moves the helix out of position during RNA binding, while other proteins bind to RNA using the loop regions[4]. RRM 3 may also be stabilized by these helices. The second project reported here was used to extend the DNA sequence (and therefore the translated amino acid sequence) of SHARP RRM 2-4 by a two-step PCR extension and Gateway cloning.

As stated above SHARP RRM 2 has been purified and used in various 2D and 3D NMR experiments. The unbound structure is currently being assigned by a graduate student in Dr. Leeper's lab. In order to determine the structure of the protein-RNA complex, NMR paramagnetic relaxation enhancement (PRE) will be used. PRE is a useful technique for large complexes due to the lack of long range

nuclear overhauser effect resonances (NOEs) as the complex increases in size. NOEs allow for distances of 6 Angstroms to be measured while PREs allow distances up to 35 Angstroms. This is due to the addition of a spin label (here a nitroxide radical) causing a magnetic dipole interaction with the unpaired electron and a nuclear relaxation rate increase[5]. The spin-labeled samples are then used in 2D heteronuclear single quantum coherence spectroscopy (HSQC) experiments with ^{15}N labeled protein. These distances can show exactly which residues are interacting with the RNA during binding. The final project reported here was used to mutate and label SHARP RRM 2 for use in NMR PRE experiments on the SHARP RRM 2-RNA complex.

Materials and Methods

Polymerase Chain Reaction (PCR) Extension: Phusion ® High-Fidelity PCR Kit (New England Biolabs) was used for SHARP RRM 2-4 sequence extension by using the primers SHARP 2-4 N-terminal, C-terminal 1, and C-terminal 2 (Appendix 3, Integrated DNA Technologies). Thermal cycler programs for each extension reaction can be found in Appendix 3. PCR reaction products were run on a 1% agarose gel (100 V, ~90 minutes for efficient separation) and inspected with UV light to determine amplicon size (in base pairs). Second extension sequence was gel extracted and purified using an Accuprep ® Gel Purification Kit (Bioneer). Samples were sent to GeneWiz Inc. for Sanger DNA sequencing using primers SHARP 2-4 N-terminal and C-terminal 2.

Gateway Cloning: SHARP RRM 2-4 PCR extension products were used in a Gateway ® Technology Cloning Kit (Invitrogen by Life Technologies). The BP reaction product was transformed into DH5-α competent cells. E.Z.N. A. Plasmid DNA Mini Kit II (Omega) spin protocol was used to purify the entry clone plasmid pDONR 221 from the colonies. The entry clone plasmid was used in the LR reaction to insert the SHARP RRM 2-4 sequence into an MBP-fusion vector, pDEST-HisMBP. The LR reaction product was transformed into One Shot ® OmniMAX ^{TM} 2 T1 Phage-Resistant *E. coli* cells. E.Z.N. A. Plasmid DNA Mini Kit II (Omega) spin protocol was used to purify the cloned plasmid from the colonies. Samples were sent to GeneWiz Inc. for Sanger DNA sequencing using primers MBP open reading frame (ORF), MBP forward insert, and MBP reverse insert (Appendix 3). Control BP reaction

was done using pEXP7-tet positive control. Product plasmid was transformed into DH5-α and plated on LB-Ampicillin plates soaked with 20 mg/ml tetracycline.

Site Directed Mutagenesis: SHARP RRM 2 point mutations were produced using Quikchange II Site Directed Mutagenesis Kit (Agilent Technologies) and primers S26C1, S26C2, S50C1, and S50C2 (Appendix 3, Integrated DNA Technologies). Thermal cycler program for the reaction can be found (Appendix). The resulting plasmid product was transformed by heat shock into XL-10 Gold competent cells and plated on LB-Ampicillin. E.Z.N. A. Plasmid DNA Mini Kit II (Omega) spin protocol was used to purify the mutant plasmid from colonies. Samples sent to GeneWiz Inc. for Sanger DNA sequencing using TEV cleavage sequence primer (Appendix 3).

Protein Expression and Purification: SHARP RRM 2 plasmids were transformed by heat shock into competent *Escherichia coli* strain BL21-DE3, a strain suitable for protein expression. After a 30 minute incubation (37°C, 260 rpm) in 500 ul SOC media the cells were plated on LB-Ampicillin agar plates and incubated at 37°C overnight. One colony from a plate was transferred to 50 ml of LB-Ampicillin media and grown overnight (37°C, 260 rpm, 12-16 hours). When an isotopic label was required, a minimal M9 media was used with $^{15}NH_4Cl$ as the sole nitrogen source. From the small grow up, 25 ml was transferred to 500 ml of LB-Ampicillin media and grown at 37°C until the absorbance at 600 nm was 0.4-0.6 using the LB-Ampicillin media as a blank. The bacteria was induced with 50 uM isopropyl beta-D-1-thiogalactopyranoside (IPTG) and grown overnight (18°C, 260 rpm, 12-16 hours). The media was centrifuged and the supernatant media was discarded. The pellets were resuspended in a buffer containing 200 mM KCl, 20 mM KPhos pH 7.6, 2 mM DTT, and 20 mM imidazole and lysed using a French Pressure Cell. The resulting solution was centrifuged and the supernatant media was saved. Each protein contains a 6-histadine tag allowing for the performance of immobilized metal affinity chromatography (IMAC) on the protein of interest. IMAC was run using an AKTApurifier and Frac-920 (GE Healthcare Life Sciences). The 6-His tag binds to the nickel ions in the column and the protein was eluted using a second buffer containing 200 mM KCl, 20 mM KPhos pH 7.6, 2 mM DTT, and 400 mM imidazole. The eluted protein was collected and combined for dialysis in a buffer containing 200 mM KCl, 20 mM KPhos pH 7.6, and 2 mM DTT. All of the SHARP RRM plasmids also contain a tobacco etch virus (TEV) cleavage site immediately before the protein of interest. The protein was cleaved

from the tag using 1 OD_{280} of TEV protease/100 OD_{280} of protein after one hour of dialysis. The cleavage and dialysis proceeds overnight (4°C). After cleavage, a second IMAC column was performed to isolate the protein of interest from the tag. The eluted protein was collected and dialyzed overnight (4°C) to remove any residual imidazole. The dialyzed protein solution was concentrated to ~2 ml using an Amicon stirred cell under 60 psi nitrogen gas.

Site Directed Spin Labeling (SDSL): Mutant SHARP RRM 2 proteins were reacted with a 10 molar excess of (1-Oxyl-2,2,5,5-tetramethyl-delta3-pyrroline-3-methyl) Methanethiosulfonate (MTSL, Cayman Chemical) for 12-16 hours at 4°C[6]. Unreacted MTSL was removed by size exclusion chromatography (SEC) using a buffer containing 80 mM KCl, 40 mM KPhos pH 7.6, and 2 mM DTT on an AKTAprimeplus and HiLoad 16/60 Superdex 75 prep grade SEC column (GE Healthcare Life Sciences). The eluted protein was collected and concentrated to ~1 ml using an Amicon stirred cell under 60 psi nitrogen gas.

^{15}N HSQC NMR experiments: Protein samples (0.2-0.5 mM) were dialyzed against an NMR buffer containing 80 mM KCl, 40 mM KPhos pH 7.6, and 2 mM DTT to remove excess salt. NMR samples (500 ul) contained 8% deuterium oxide (D_2O). ^{15}N HSQC spectra were collected using vNMRj on a 750 MHz Varian NMR spectrometer. Spectra were processed using NMRpipe to phase, base line correct, and zero fill (size 8192). Assignments were made by using Sparky (NMRfam) by comparing the mutant spectra to the wild type spectrum. The same experiments were run on the spin labeled samples.

Results and Discussion

SHARP RRM 3 Purification Optimization: Multiple grow ups of SHARP RRM 3 were produced by using the Protein Expression and Purification procedure described above. See Appendix 2: Table 1 for each buffer condition. Low salt and salt change exhibited protein aggregate during the first dialysis step before cleavage of the 6-His tag. The protein purified using high salt and high reductant buffers aggregated overnight after the first dialysis step and cleavage. Addition of arginine in the buffer allowed the protein to be concentrated to 57 uM without aggregation, and the addition of arginine and glutamate allowed the protein to be concentrated to 60 uM. Based on Arieti et al. (2014)[4], SHARP RRM

3 is able to be purified, but in their study only supplementary circular dichroism and SEC confirmed the presence of RRM 3. RRM 3 was also found to be tightly associated with RRM 4, indicating that the C-terminal helices of RRM 4 may stabilize RRM 3. The RRM3 plasmid used by Dr. Leeper's group was larger than the RRM3 plasmid purified by Arieti et al. (2014)[4], containing six additional N-terminal residues and seven additional C-terminal residues (Appendix 2: Figure 2). Based on the crystal structure these additional residues are unstructured and include a cysteine residue. These features may have led to the constant aggregation and possible disulfide bond formation of the RRM 3 samples.

Figure 2-SHARP RRM 4 showing Dr. Leeper's original plasmid (black) and the additional C-terminal helices (red). Based on PDB 4P6Q and generated using PyMOL

SHARP RRM 2-4 PCR Extension Primer Design: Using the PDB file 4P6Q deposited by Arieti et al. (2014)[4] it was determined that the SHARP RRM 2-4 plasmid used by Dr. Leeper's lab exculed the C-terminal sequence: LAFTHCMEKSQDIRDFYEMLAERREER (residues 593-620, Figure 2). Primers for the extension were designed so that the DNA sequence would be codon optimized for *E. coli* based on the SHARP RRM 2-4 plasmid sequence. This technique minimizes RNA secondary structure during transcription and translation, and limits G-rich sequences. These steps are taken so that the ribosome can efficiently produce the protein without errors caused by low concentration of tRNA for a non-optimized codon or stopping due to secondary structure of the RNA. Trial and error of the codon combinations produced the least G-C bonds and the lowest energy RNA secondary structure (Figure 3a). Gateway cloning also uses specific sequences on the 5' and 3' ends that act as 5' sticky ends during each reaction (attB flanking sites)[7]. The extended SHARP 2-4 sequence and the Gateway sequences were used as a template for the primers SHARP 2-4 N-terminal, C-terminal 1 and C-terminal 2 (Appen-

dix 3, Integrated DNA Technologies). SHARP 2-4 N-terminal and C-terminal 1 have an overlap region with the original plasmid, while C-terminal 2 has an overlap region with C-terminal 1.

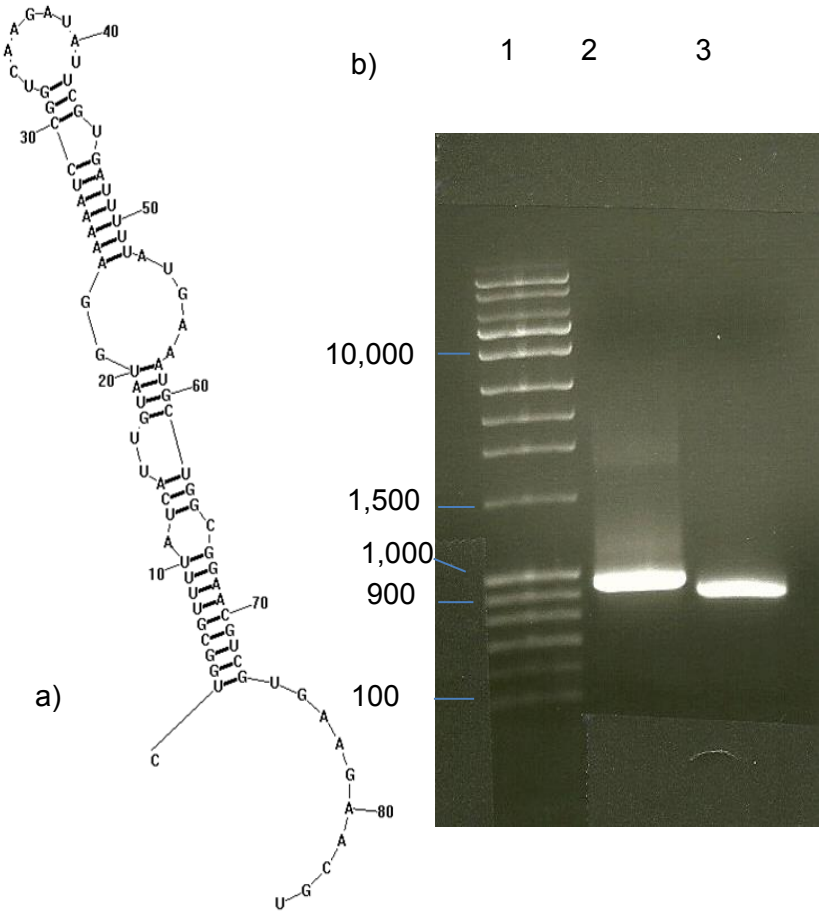


Figure 3- (a) Secondary structure RNA model for codons of the extended SHARP RRM 2-4 sequence, produced using RNAstructure 5.6. The secondary structure is unlikely to begin until approximately halfway through the sequence and limits the amount of potential G-C bonding. (b) Using primers based on this sequence, the resulting PCR extension products were run on an agarose gel. Lane 1 contains the ladder, Lane 2 contains extension two, and Lane 3 contains extension one.

SHARP RRM 2-4 PCR Extension Reaction: Due to the length of the additional sequence, the PCR was split into two reactions (Appendix 2: Figure 3). The first reaction used SHARP RRM 2-4 plasmid and primers SHARP 2-4 N-terminal and C-terminal 1. Different reaction cycles were determined for each reaction step based on the primer overlap melting temperature and the overall primer melting temperature. Three cycles were used for ambient annealing, three for overlap annealing, and the rest for full sequence annealing. The thermal cycler program for this reaction can be found in Ap-

pendix 3: Table 2. The first reaction should produce an amplicon of 925 base pairs. The second reaction used the first PCR reaction product and primers SHARP 2-4 N-terminal and C-terminal 2. The thermal cycler program for this reaction can be found in Appendix 3: Table 3. The second reaction should produce an amplicon of 983 base pairs. Both reactions were run on the 1% agarose gel (Figure 3b). Based on this gel the second extension reaction creates an amplicon larger than the first extension and the product was used in subsequent experiments.

SHARP RRM 2-4 Gateway Cloning Reactions: Gateway cloning consists of two reactions, beginning with the BP reaction which inserts the attB flanked PCR product into an entry clone vector pDONR-211. From the entry clone vector there are multiple destination vectors that can be purchased for use in the LR reaction. Each reaction uses an enzyme mix containing integrase, excisionase, and ligase which allow the exit and entry of the sequence of interest between plasmids. The SHARP RRM 2-4 sequence was integrated into the pDEST-HisMBP vector. The pDONR vector and pDEST vectors contain the gene ccdB which inhibits the growth of *E. coli* and is removed during the reaction. This is to ensure that only the plasmids that contain the gene of interest replicate in the colonies. The BP reaction was incubated at room temperature for 4 hours and frozen (-20°C) until the second (LR) reaction. The product of the LR reaction was transformed into One Shot ® Omnimax ™ 2 T1 Phage-Resistant *E. coli* cells. Plasmid isolation of three colonies were sent to GeneWiz Inc. for Sanger DNA sequencing using primers for MBP ORF (bp 1446-1469), MBP forward (bp 2620-2643), and MBP reverse (bp 4413-4437, Figure 4). Results from this reaction showed that the MBP forward and reverse had no priming, while the MBP ORF sequence was still in the plasmid. Based on this observation a positive control BP reaction was performed. The BP product plasmid with the control pEXP7-tet gene should produce >100 DH5-α colonies on LB plates soaked in 20 mg/ml tetracycline. However, the control reaction only produced eight colonies. The existence of colonies with the MBP ORF and the failure of the positive control led to the conclusion that part of the enzyme mix had been degraded since last used by members of Dr. Leeper's lab. If colonies survive, but don't contain the gene of interest, it is likely that the integrase is degraded. Excisionase and ligase must be active as the ccdB suppressor gene was removed from the plasmid for colony survival.

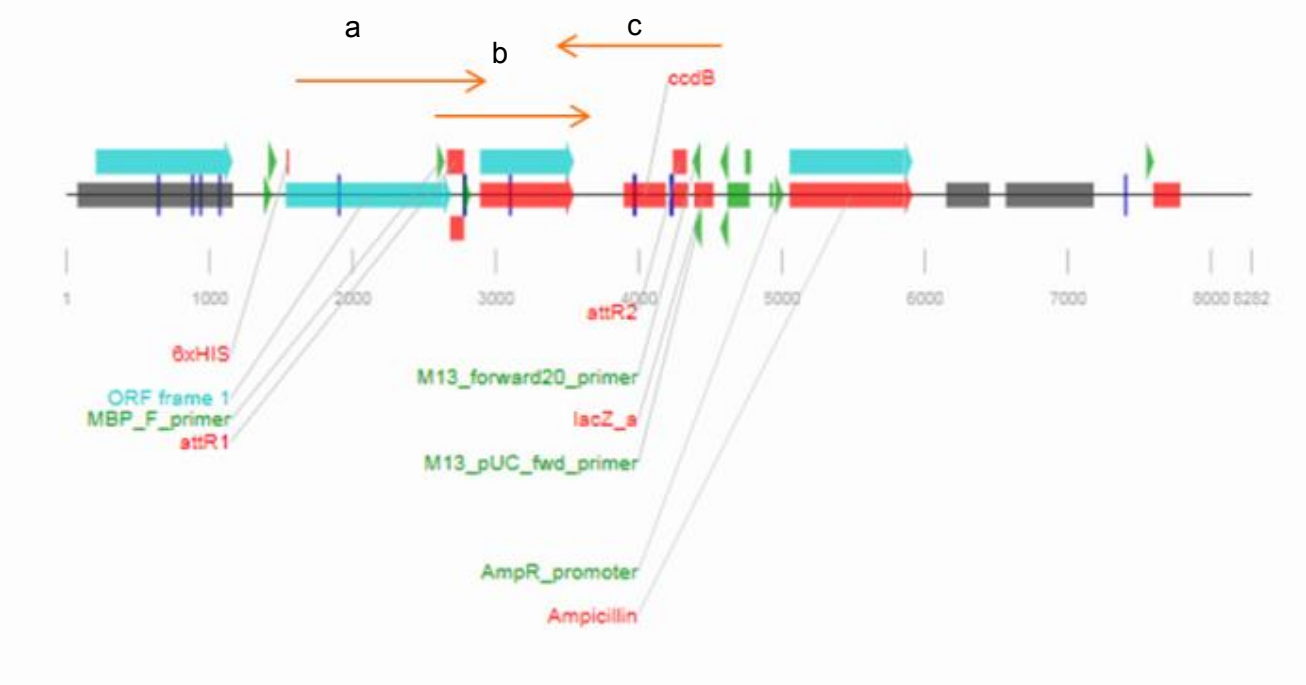


Figure 4-Primer locations for Sanger sequencing of MBP-fusion vector with SHARP RRM 2-4 insert. Three primers were used in each sample MBP ORF (a), MBP forward insert (b), and MBP reverse insert (c). attB flanking sites allow the protein of interest to be inserted between sites b (attR1) and c (attR2). Gene figure modified from www.addgene.org/browse/sequence/6897.

SHARP RRM 2-4 PCR Extension Reaction Sequencing: A second two step extension of the SHARP RRM 2-4 plasmid was performed and run on a 1% agarose gel. The 983 base pair band was excised from the gel and purified using Accuprep Gel Purification Kit (Bioneer). Appendix 2: Figure 4 shows the absorption of the purified product. There is a large absorbance at 240 nm, assumed to be salt from the elution buffer of the last purification step; however, the 260 nm reading shows 51 ng DNA/ul. The sample was sent to Genewiz Inc. for Sanger sequencing using the SHARP 2-4 N-terminal and C-terminal 2 primers. The quality score for both reactions was too low to pass the quality standard, but usable sequence data was obtained from the C-terminal 2 sequence. This sequence was translated and aligned with the proposed amino acid sequence for SHARP RRM 2-4 with the additional helices for an 85% match in 221 residues sequenced with the extended sequence appearing at the C-terminal end of the protein (Appendix 2: Figure 5). If the group finds it necessary to sequence at the 5' and 3' end of the construct it is possible that internal primers will be needed, rather than the 5' and 3' primers used in these sequencing reactions. Based on these reactions and the distinct amplicon size change seen in the agarose gel (Figure 3b) the sequence is assumed to be extended as intended.

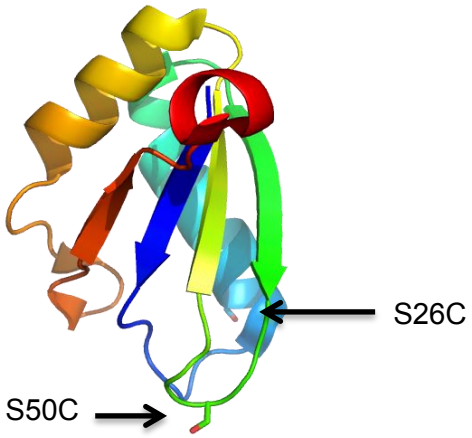


Figure 5-S26C and S50C locations in SHARP RRM 2 structure, generated using PyMOL.

SHARP RRM 2 Site Directed Mutagenesis: In order to attach a spin label to a protein for PRE there must be a single specific site for the label to bind. No cysteine residues occur in the SHARP 2 structure, but these can be introduced using a point mutation of a serine residue codon. Eight serine residues occur in the SHARP RRM 2 structure, but the mutation must be chosen carefully for correct labeling. Residues must have a high solvent accessibility to ensure that alkylation occurs. The residue must not be in an area of structural stability or activity to avoid any difference in the protein folding and binding once the label is bound[8]. The mutants for SHARP RRM 2, S26C and S50C, can be seen in Figure 5. S26C occurs near the end of one of the alpha helices and S50C occurs in a variable loop region. Each residue is on a separate face of the protein, but near to each other in space. Point mutations were made using a Quikchange II Site Directed Mutagenesis Kit (Agilent Technologies) and the primers S26C1, S26C2, S50C1 and S50C2 (Appendix 3). The thermal cycler program for this reaction can be found in Appendix 3: Table 4. The reaction plasmid was transformed into XL-10 Gold competent cells and plated on LB-Ampicillin agar plates. Four plasmid isolations for each mutation (350 ul) were combined and spin-vacuumed for two hours to dehydrate the sample and the resulting DNA pellet was resuspended in 50 ul of sterile water. Each mutation plasmid was sent to Genewiz Inc. for Sanger sequencing with the TEV sequence primer (Appendix 3). The resulting sequences were translated and aligned to the wild type sequence (Appendix 2: Figure 6). Both mutations occurred at the correct residue and did not affect other residues. Protein samples were prepared as stated in the Materials and Methods by using minimal M9 media with $^{15}NH_4Cl$ as the sole nitrogen source.

SHARP RRM 2 Site Directed Spin Labeling: An aliquot of 1 ml of the isotopically labeled and concentrated protein was reacted with a 10 molar excess of MTSL from a 200 mM stock in 100% acetonitrile. MTSL is similar in size and in molecular volume to a tryptophan residue and should have little to no effect on structure and stability of the sample[9]. SHARP RRM 2 S26C (0.364 mM) was reacted with 3.64 mM of MTSL solution. SHARP RRM 2 S50C (0.329 mM) was reacted with 3.29 mM MTSL solution. Each alkylation reaction (Figure 6) was incubated at 4°C for 12-16 hours and immediately transferred to an AKTAprimeplus and HiLoad 16/60 Superdex 75 prep grade SEC column (GE Healthcare Life Sciences) using the NMR buffer to remove unreacted MTSL. The eluted protein was collected and concentrated to ~1 ml. SHARP RRM 2 S26C had aggregation during SEC and was concentrated to 0.167 mM. SHARP RRM 2 S50C was concentrated to 0.217 mM.

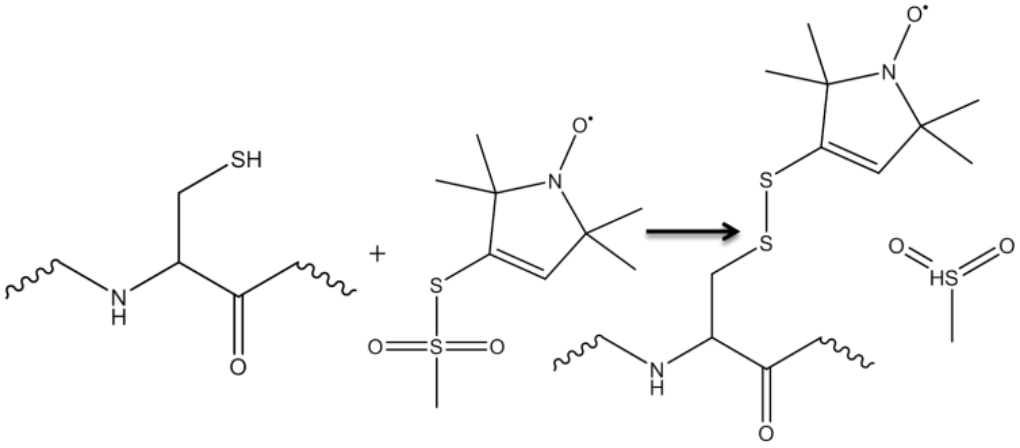


Figure 6-Reaction of cysteine residue the nitroxide spin label, generated using ChemDraw 13.0. Unreacted MTSL and the resulting by-product were separated from the protein by SEC.

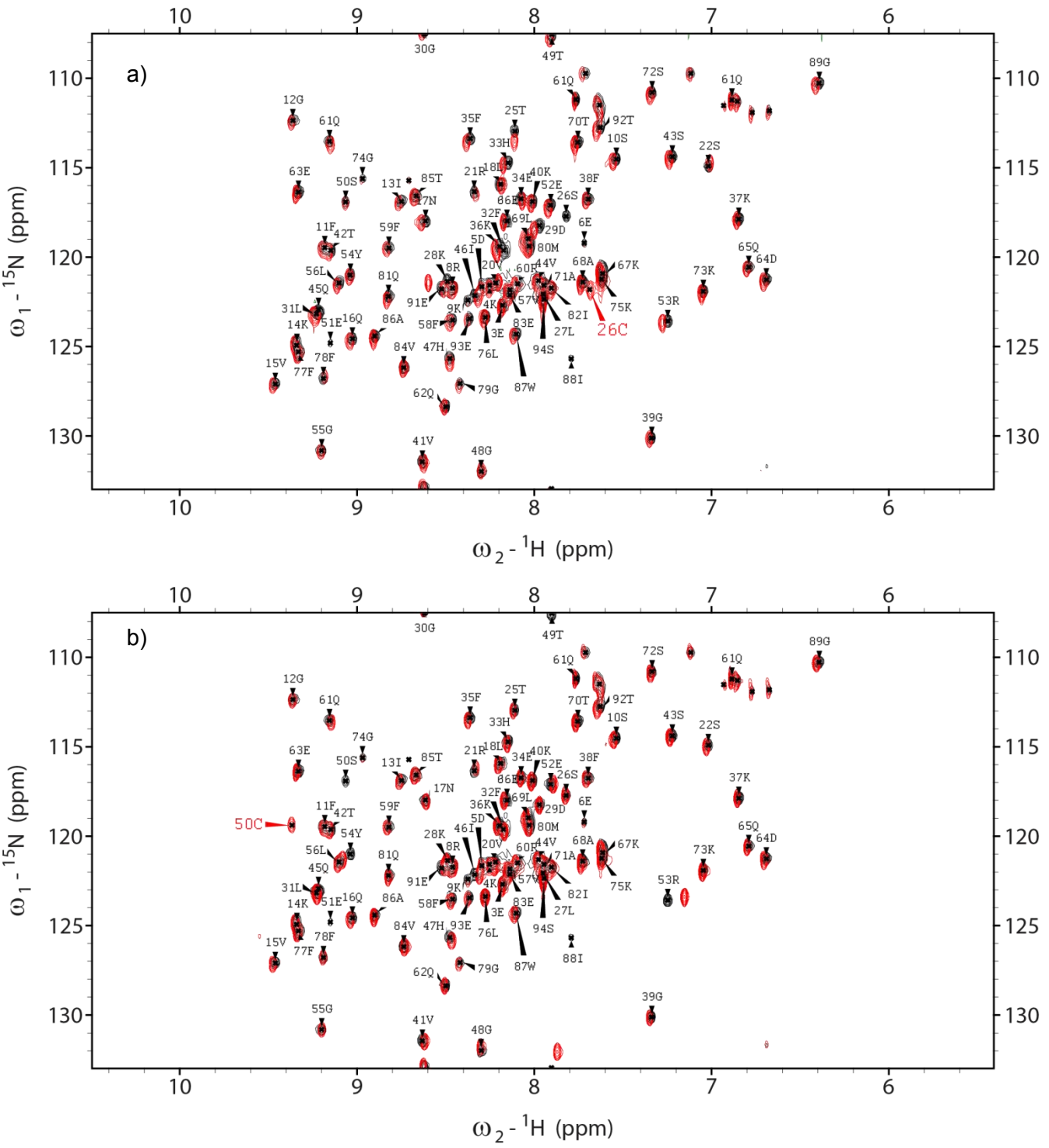


Figure 7- ^{15}N HSQC data for mutants samples before MTSL addition with wild type spectra in black and mutant spectra in red. (A) shows SHARP RRM 2 S26C spectrum and (b) shows SHARP RRM 2 S50C spectrum. The SHARP RRM 2 ^{15}N HSQC NMR section of Results and Discussion explains where resonances were shifted. Figures generated by using Sparky (NMRfam).

SHARP RRM 2 ^{15}N HSQC NMR: Protein NMR samples (500 ul) contained 8% D_2O. S26C and S50C samples were run before adding spin label to ensure that most resonances were not shifted and

new residue assignments would not be necessary. SHARP RRM 2 S26C non spin-labeled HSQC showed shifts in 25T, 27L, 28K, and a new resonance, which does not appear in the wild type spectrum, that may be 24D (Figure 7a). The resonance that is shifted in the nitrogen dimension from the wild type 26S resonance is presumed to be 26C. SHARP RRM 2 S50C non spin-labeled HSQC shows shifts in 46I, 49T, 53R, and 54Y (Figure 7b). The resonance that is shifted in the nitrogen dimension from the wild type 50S resonance is presumed to be 50C. S26C and S50C samples were also run after the spin labeling reaction to observe the broadening of peaks due to PRE and as a comparison for the RNA bound spectra. Spin labeled samples were run using the same procedure, however, the concentration of both samples was lower after the labeling reaction. SHARP RRM 2 S26C spin labeled HSQC showed loss of resonance for residues 21R, 22S, 25T, 50S, 55G,78F, 79G, and 93E (data not shown). SHARP RRM 2 S50C spin labeled HSQC showed loss of resonance for residues 6E, 9K, 74G, 78F, and 93E (data not shown).

Conclusions

Determining the structure of the unbound and RNA bound RRMs is necessary to understand how the interactions occur and change conformation of both the RNA and protein affects the signaling pathways. These research projects were focused on preparing the truncated RRMs for 2D and 3D NMR experiments. RRM 3 was difficult to purify using the procedure used for SHARP RRM 1 and 2. The optimization of the RRM was relatively unsuccessful, but the addition of arginine and glutamate did allow a small concentration of protein to survive the purification process. However, recent research has shown that RRM 3 and RRM 4 are tightly associated and the C-terminal helices of RRM 4 may stabilize the protein[4]. Based on this research, focus was shifted to obtaining a SHARP RRM 2-4 plasmid that contained the missing C-terminal helices, by using PCR extension and Gateway cloning. The PCR reactions to extend the plasmid are assumed to have worked based on agarose gel and partial sequencing data. Research on RRM 2-4 will continue once a new Gateway ® Technology Cloning Kit (Invitrogen by Life Technologies) is purchased by adopting the primers and techniques discussed in this report. The protein purification will be attempted using the buffer system in Arieti et al. (2014)[4], and used in various 2D and 3D NMR experiments to determine the unbound and RNA bound structures. RRM 2 cysteine mutations for NMR PRE were successful and based on the non spin-labeled 15N HSQC spec-

tra for S26C and S50, do not affect the protein overall structure. Spectra of the MTSL labeled protein show loss of some resonances, but this may be due to the lower concentration of the samples obtained after the labeling reaction. The current experiment focuses on increased T2 relaxation times to determine if ^{15}N HSQC resonances near the spin label broaden faster than those at a larger distance. Further analysis can be done by using mass spectrometry to determine the molecular weight of the spin labeled samples. MTSL should increase the molecular weight of the protein by 185 g/mol. One of the unsolved problems with the NMR experiments is the optimization of salt concentration, which could increase resolution if the salt concentration is decreased. Further research with the cysteine mutants will include RNA-bound complex structures and include analysis of the spin labeled protein with unlabeled RNA and unlabeled protein with spin labeled RNA. This will require structure assignments for each RNA fragment used which has yet to be completed by Dr. Leeper's group. However, the cysteine mutants can also be fluorescently labeled with thiol-reactive dyes and used in RNA footprinting experiments without assignments for the RNA structure. Electron paramagnetic resonance (EPR) and double electron-electron resonance (DEER) spectroscopy with nitroxide spin labeled cysteine mutants can also be performed for additional complex structure determination methods[10,11]. Cysteine mutants will also be made for SHARP RRM 1, which does not contain cysteine residues and has four potential serine residues for mutation. The same labeling experiments can then be run using with RRM 1 and the associated RNA fragments.

Acknowledgements

Training for protein expression and purification, running and processing of NMR experiments, and previous SHARP RRM data were provided by graduate student Joel Caporoso. Assistance with Gateway cloning primer design was provided by graduate student Dan Morris. SHARP RRM 1 structure for serine residue locations (figure not shown) provided by graduate student Caroline Davis. Funding and equipment provided by Dr. Thomas Leeper and the Department of Chemistry at the University of Akron.

References

[1] Shi, Y.; Downes, M.; Xie, W.; Kao, H.Y.; Ordentlich, P.; Tsai, C.C.; Hon, M.; Evans, R.M. Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev*. **2001**, *15*, 1140-1151.

[2] Gräslund, S.; Nordlund, P.; Weigelt, J.; Bray, J.; Gileadi, O.; Knapp, S.; Oppermann, U.; Arrowsmith, C.; Hui, R.; Ming, J.; Dhe-Paganon, S.; Park, H.W.; Savchenk, A.; Yee, A.; Edwards, A.; Vincentelli, R.; Cambillau, C.; Kim, R.; Kim, S.H.; Rao, Z.; Shi, Y.; Terwilliger, T.C.; Kim, C.Y.; Hung, L.W.; Waldo, G.S.; Peleg, Y.; Albeck, S.; Unger, T.; Dym, O.; Prilusky, J.; Sussman, J.L.; Stevens, R.C.; Lesley, S.A.; Wilson, I.A.; Joachimiak, A.; Collart, F.; Dementieva, I.; Donnoelly, M.I.; Eschenfeldt, W.H.; Kim, Y.; Stols, L.; Wu, R.; Zhou, M.; Burley, S.K.; Emtage, J.S.; Sauder, J.M.; Thompson, D.; Bain, K.; Luz, J.; Gheyi, T.; Zhang, F.; Atwell, S.; Almo S.C.; Bonanno, J.B.; Fiser, A.; Swaminathan, S.; Stuider, F.W.; Chance, M.R.; Sali, A.; Acton, T.B.; Xiao, R.; Zhao, L.; Ma, L.C.; Hunt, J.F.; Tong, L.; Cunningham, K.; Inouye, M.; Anderson, S.; Janjua, H.; Shastry, R.; Ho, C.K.; Wang, D.; Wang, H.; Jiang, M.; Montelione, G.T.; Stuart, D.I.; Owens, R.J.; Daenke, S.; Schutz, A.; Heinemann, U.; Yokoyama, S.; Bussow, K.; Gunsalus, K.C. Protein production and purification. *Nat. Methods* **2008***, 5*, 135-146.

[3] Golovanov, A.P.; Hautbergue, G.M.; Wilson, S.A.; Lian, L.Y. A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc*. **2004**, *126*, 8933– 8939.

[4] Arieti, F.; Gabus, C.; Tambalo, M.; Huet, T.; Round, A.; Thore, S. The crystal structure of the Split End protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. *Nucleic Acids Res.* **2014**, *42*, 6742-6752.

[5] Clore, G.; Iwahara, J. Theory, Practice and Applications of Paramagnetic Resonance Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* **2009**, *109*, 4108-4139.

[6] Liang, B.; Bushweller, J.; Tamm, L. Site-Directed Parallel Spin Labeling and Paramagnetic Relaxation Enhancement in Structure Determination of Membrane proteins by solution NMR spectroscopy. *J. Am. Chem. Soc.* **2006**, *128*, 4389-4397

[7] Nallamsetty, S.; Waugh, D.S. A generic protocol for the expression and purification of recombinant proteins in *Escherichia coli* using a combinatorial His_6-maltose binding protein fusion tag. *Nat. Protoc.* **2007**, *2*, 383-391

[8] Gillespie, J.; Shortle, D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **1997**, *268*, 170-184

[9] Czogalla, A.; Pieciul, A.; Jezierski, A.; Sikorski, A. Attaching a spin to a protein—site-directed spin labeling in structural biology. *Acta Biochim. Pol. (Engl. Transl.)* **2007**, *54*, 235-244

[10] Edwards, T.; Robinson, B.; Sigurdsson, S. Identification of amino acids that promote specific and rigid TAR RNA-Tat Protein Complex Formation. *Chem. Biol.* **2005**, *12*, 329-337

[11] Altenbach, C.; Kusnetzow, A.; Ernst, O.; Hofmann, K.; Hubbell, W. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7439-7444

Appendix 1: Safety Information

Equipment Training-Most equipment training was received from University of Akron graduate student Joel Caporoso. Potential hazardous equipment included, but not limited to; French pressure cell (which uses up to 1200 psi to lyse cells), high speed centrifuge (if incorrectly balanced), and NMR magnets (which can pull in loose metallic objects). Autoclave training was provided by Dr. Hazel Barton of the University of Akron Biology Department.

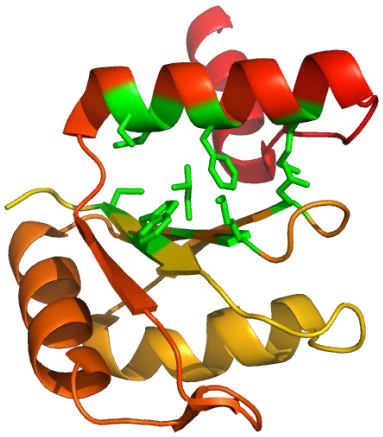
Chemical safety-Gloves and googles were worn when working with any chemical substances. Sulfur containing solutions such as DTT, BME, and MTSL were used in the chemical hood to prevent eye and nasal irritation and colleagues were alerted of the chemical being used. Pipette tips are disposed of in cardboard or Styrofoam boxes. Syringe needles are disposed of in metal containers to avoid punctures that would occur if disposed in plastic or cardboard. Buffer systems used in IMAC FPLC and NMR were disposed of in the sink with copious amounts of water. Ethinium bromide, used in DNA-binding in agarose gel, was added to the gel solution and the flasks were washed immediately. Agarose gels are disposed of by allowing them to dry out. Halogen and non-halogen waste containers are located in the flammable solvent cabinet and disposed of by the Department of Environmental and Occupational Health and Safety.

Biohazard safety-Gloves and googles were worn when working with any biological substances. Pipette tips that come in contact with bacteria or bacterial media are disposed of in a separate tip box than chemical tips. LB-agar plates that have been used for bacteria growth were disposed of in a specific biohazard waste container. Waste containers are disposed of by the Department of Environmental and Occupational Health and Safety. Bacterial media is diluted with bleach in the bottle or flask after use to deter any further growth and disposed of in the sink with copious amounts of water.

NMR Lab safety-The magnetic field of the 750 MHz NMR exerts a large force and may pull loose metallic objects, due to the NMR remaining on when not running an experimental sample.

The physical magnet has a large current running through it and is cooled by large amounts of liquid nitrogen and helium, which will evaporate if a leak occurs in the shielding. Liquid helium is approximately -269°C and liquid nitrogen is approximately -196°C.

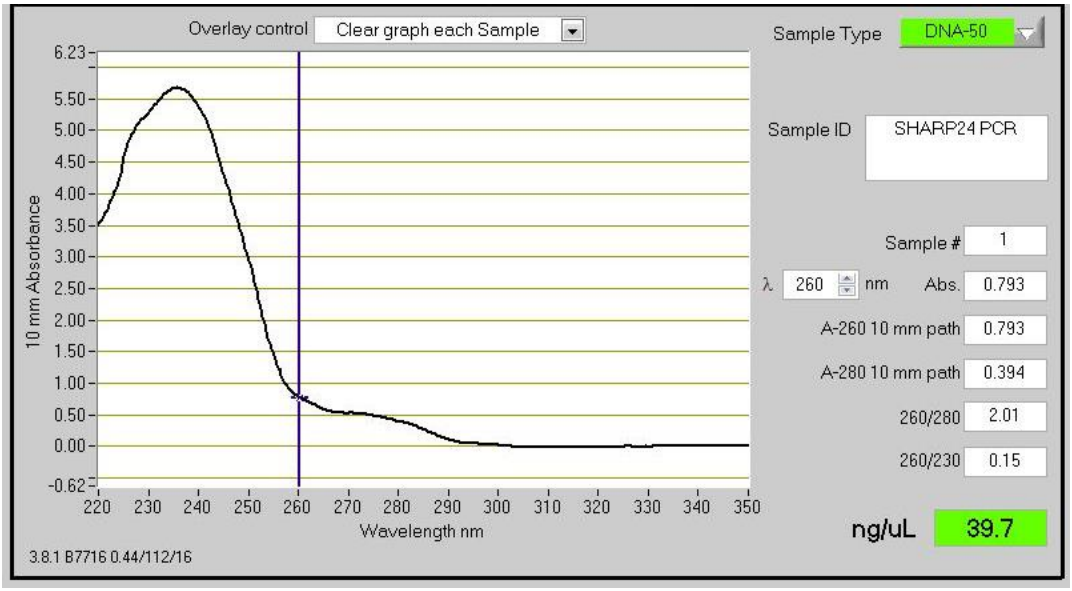
Appendix 2-Supplementary Figures



Supplementary Figure 1-Rotated view of SHARP RRM 4 to show the residues (green) which may interact between the C-terminal helix and the beta sheet surface. Generated using PyMOL based on PDB 4P6Q.

Supplementary Figure 2-SHARP RRM 3 (black) with the unstructured regions (red) that could decrease stability and were not present in the Arieti et al. (2014) truncated plasmid.

Supplementary Figure 3- Two step PCR extension of SHARP RRM 2-4 DNA sequence. a). Reaction one using primer N-terminal and C-terminal 1. attB1 flanking site is added to the 5' end of the sequence, overlapping with the TEV protease cleavage site. Red sequence indicates codons that are part of the C-terminal alpha helices. b). Reaction two using primers N-terminal and C-terminal two. attB2 flanking site is added to the 3' end of the sequence overlapping with the extended sequence from reaction one.



Supplementary Figure 4-Nanodrop absorbance curve of the gel purified SHARP RRM 2-4 PCR extension two product. The large absorbance at 230-240 nm is most likely salt from the purification kit's elution buffer.

Range 1: 25 to 244 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
374 bits(959)	2e-134	Compositional matrix adjust.	187/221(85%)	189/221(85%)	2/221(0%)

```
Query  60   RQQEDQEKALTASKGKLFFGMQIEVTAWIGPETESENEFRPLDERIDEFHPKATRTLFIG  119
            RQQ D   AL ASK   F  +         P     N+ RP DERI EFHPKATRTLF G
Sbjct  25   RQQVDXINALDASKSTXF-SICXXXLXVDHPGNXXXNDXRPXDERIXEFHPKATRTLFXG  83

Query  120  NLEKTTTYHDLRNIFQRFGEIVDIDIKKVN-VPQYAFLQYCDIASVCKAIKKMDGEYLGN  178
            NLEKTTTY DLRNIFQRFGEIVDIDIKKVN VPQYAFLQYCDIASVCKAIKKMDGEYLGN
Sbjct  84   NLEKTTTYXDLRNIFQRFGEIVDIDIKKVNGVPQYAFLQYCDIASVCKAIKKMDGEYLGN  143

Query  179  NRLKLGFGKSMPTNCVWLDGLSSNVSDQYLTRHFCRYGPVVKVVFDRLKGMALVLYNEIE  238
            NRLKLGFGKSMPTNCVWLDGLSSNVSDQYLTRHFCRYGPVVKVVFDRLKGMALVLYNEIE
Sbjct  144  NRLKLGFGKSMPTNCVWLDGLSSNVSDQYLTRHFCRYGPVVKVVFDRLKGMALVLYNEIE  203

Query  239  YAQAAVKETKGRKIGGNKIKVDFANRESQLAFYHCMEKSGQ  279
            YAQAAVKETKGRKIGGNKIKVDFANRESQ  FYHCMEKSGQ
Sbjct  204  YAQAAVKETKGRKIGGNKIKVDFANRESQXXFYHCMEKSGQ  244
```

Supplementary Figure 5- BLAST alignment of SHARP RRM 2-4 proposed sequence (Query) and the translated C-terminal 2 Sanger sequencing results (Sbjct). An 85% match occurred with two gaps present. The extended amino acid sequence begins at residue 268 on the query and 233 on the subject.

a)

Score	Expect	Method	Identities	Positives	Gaps
172 bits(436)	1e-61	Compositional matrix adjust.	84/85(99%)	84/85(98%)	0/85(0%)

```
Query  7   SFGIKVQNLPVRSTDTCLKDGLFHEFKKFGKVTSVQIHGTSEERYGLVFFRQQEDQEKAL  66
           SFGIKVQNLPVRSTDT LKDGLFHEFKKFGKVTSVQIHGTSEERYGLVFFRQQEDQEKAL
Sbjct  10  SFGIKVQNLPVRSTDTSLKDGLFHEFKKFGKVTSVQIHGTSEERYGLVFFRQQEDQEKAL  69

Query  67  TASKGKLFFGMQIEVTAWIGPETES  91
           TASKGKLFFGMQIEVTAWIGPETES
Sbjct  70  TASKGKLFFGMQIEVTAWIGPETES  94
```

b)

Score	Expect	Method	Identities	Positives	Gaps
172 bits(436)	1e-61	Compositional matrix adjust.	84/85(99%)	84/85(98%)	0/85(0%)

```
Query  6   SFGIKVQNLPVRSTDTSLKDGLFHEFKKFGKVTSVQIHGTCEERYGLVFFRQQEDQEKAL  65
           SFGIKVQNLPVRSTDTSLKDGLFHEFKKFGKVTSVQIHGT EERYGLVFFRQQEDQEKAL
Sbjct  10  SFGIKVQNLPVRSTDTSLKDGLFHEFKKFGKVTSVQIHGTSEERYGLVFFRQQEDQEKAL  69

Query  66  TASKGKLFFGMQIEVTAWIGPETES  90
           TASKGKLFFGMQIEVTAWIGPETES
Sbjct  70  TASKGKLFFGMQIEVTAWIGPETES  94
```

Supplementary Figure 6-BLAST Alignment of SHARP RRM 2 S26C (a) and S50C (b) with the wild type sequence (Query) and the translated Sanger sequencing results for S26C and S50C (Sbjct). Both contain the intended mutation with no effect on the surrounding residues.

Buffer	Components	Final SHARP RRM 3 Concentration
RRM 3 Original Buffer	200 mM KCl, 20 mM KPhos (pH 6), 2 mM DTT	<0.1 mM in ~0.5 ml
High Reductant	200 mM KCl, 20 mM KPhos (pH 6), 4 mM DTT	Aggregated after dialysis and cleavage
Salt Change	200 mM NaCl, 20 mM NaPhos (pH 6), 2 mM DTT	Aggregated during cleavage
High Salt	400 mM NaCl, 20 mM NaPhos (pH 6), 2 mM DTT	Aggregated after dialysis and cleavage
Low Salt	80 mM NaCl, 20 mM NaPhos (pH 6), 2 mM DTT	Aggregated during cleavage
Addition of Arginine	200 mM NaCl, 20 mM NaPhos (pH 6), 2 mM DTT, 100 mM arginine, 99% L(+)	0.0574 mM in ~4 ml
Addition of Arginine and Glutamate	200 mM NaCl, 20 mM NaPhos (pH 6), 2 mM DTT, 50 mM arginine 99% L(+), 50 mM glutamate L(+)	0.062 mM in ~1.5 ml with some aggregation

Supplementary Table 1-Buffer systems used in SHARP RRM 3 optimization experiments and the resulting protein concentration or failure.

Appendix 3: PCR Primer Sequences and Programs

SHARP 2-4 N-terminal

5’ GGG GAC AAG TTT GTA CAA AAA AGC AGG C 3’

SHARP 2-4 C-terminal 1

5’ TTC ATA AAA ATC ACG AAT ATC TTG ACC GG 3’

SHARP 2-4 C-terminal 2

5’ GGG GAC CAC TTT GTA CAA AGC TGG GTT A 3’

SHARP 2 S26C primer 1

5’ GCC GTC TTT CAG GCA CGT ATC CGT CGA A 3’

SHARP 2 S26C primer 2

5’ CGT TCG ACG GAT ACG TGC CTG AAA GAC G 3’

SHARP 2 S50C primer 1

5’ CAT AAC GTT CTT CGC AGG TGC CGT GAA TC 3’

SHARP 2 S50C primer 2

5’ CCA GAT TCA CGG CAC CTG CGA AGA ACG T 3’

TEV sequence

5’ GAA AAT CTG TAC TTC CAG 3’

MBP ORF

5’ GAG CGG ATA ACA ATT TCA CAG AGG 3’

MBP forward

5’ GAT GAA GCC CTG AAA GAC GCG CAG 3’

MBP reverse

5’ GCA AGG CGA TTA AGT TGG GTA ACG C 3’

Cycle Step	Number of Cycles	Temperature (°C)	Time (seconds)
Initial Denaturation	1	98	30
Denaturing	3	98	10
Annealing		25	15
Extension		72	20
Denaturing	3	98	10
Annealing		57.3	15
Extension		72	20
Denaturing	23	98	10
Annealing		71.6	15
Extension		72	20
Final Extension	1	72	300
Hold	1	4	

Supplementary Table 2-Thermal cycler program for PCR extension reaction one. The first three cycle period was used as an ambient annealing temperature, the second three cycle period as an overlap annealing temperature, and the twenty three cycle period as full primer annealing.

Cycle Step	Number of Cycles	Temperature (°C)	Time (seconds)
Initial Denaturation	1	98	30
Denaturing	3	98	10
Annealing		25	15
Extenision		72	20
Denaturing	3	98	10
Annealing		46.9	15
Extenision		72	20
Denaturing	23	98	10
Annealing		72.8	15
Extenision		72	20
Final Extension	1	72	300
Hold	1	4	

Supplementary Table 3-Thermal cycler program for PCR extension reaction two. The first three cycle period was used as an ambient annealing temperature, the second three cycle period as an overlap annealing temperature, and the twenty three cycle period as full primer annealing.

Cycle Step	Number of Cycles	Temperature (°C)	Time (minutes)
Initial Denaturation	1	95	0.5
Denaturing	14	95	0.5
Annealing		55	1
Extenision		68	6
Hold	1	4	

Supplementary Table 4-Thermal cycler program for Quikchange Site Directed Mutagenesis protocol. The number of cycles depends on plasmid and vector length.